
Implications of the next signaling rate on Ethernet speeds

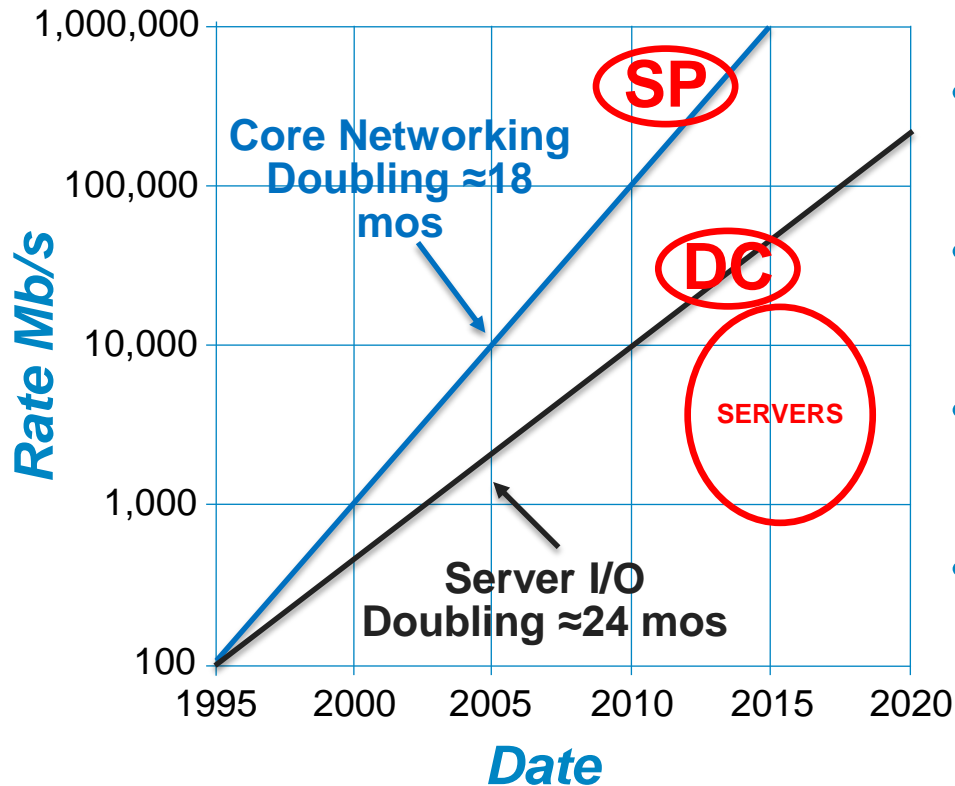


Kapil Shrikhande
Dell

Presented at the Ethernet Alliance Rate Debate TEF, October 16th, 2014

Higher Ethernet Speeds: Observations

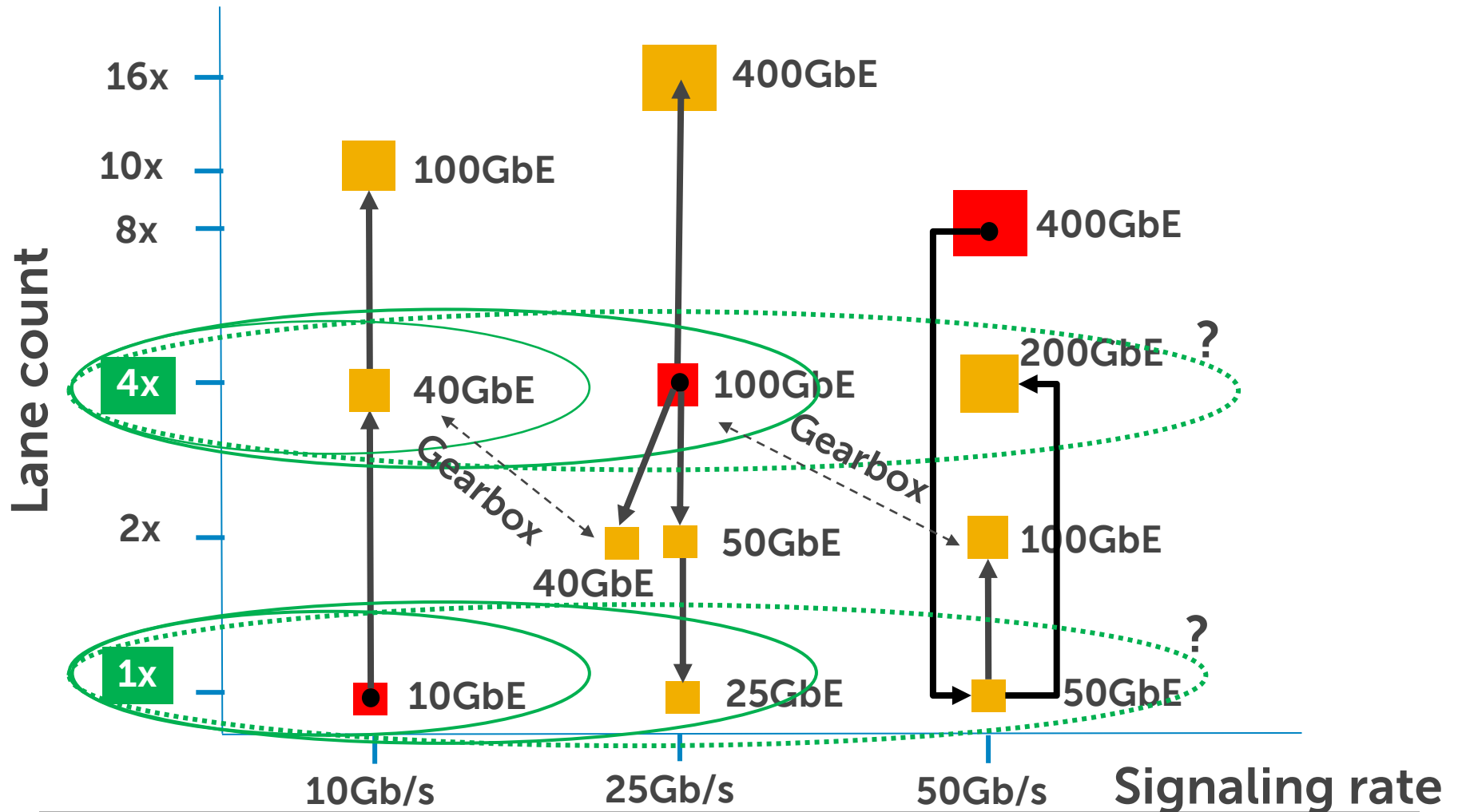
Data centers driving Ethernet differently than Core Networks



- 40G (4x10G) not 100G (10x10) took off in DC network ports
- But 100G (4x25G) will take off in DC network ports
- 25G, not 40G is likely the next volume server IO > 10G
- 400G will drive the next-gen internet core networking..
- What about Data centers? What comes after 25/100GE?
 - *Follow the serdes* 😊

Speeds, Lanes and Serdes / signaling

Data centers are building on Speeds using 1x / 4x Lanes



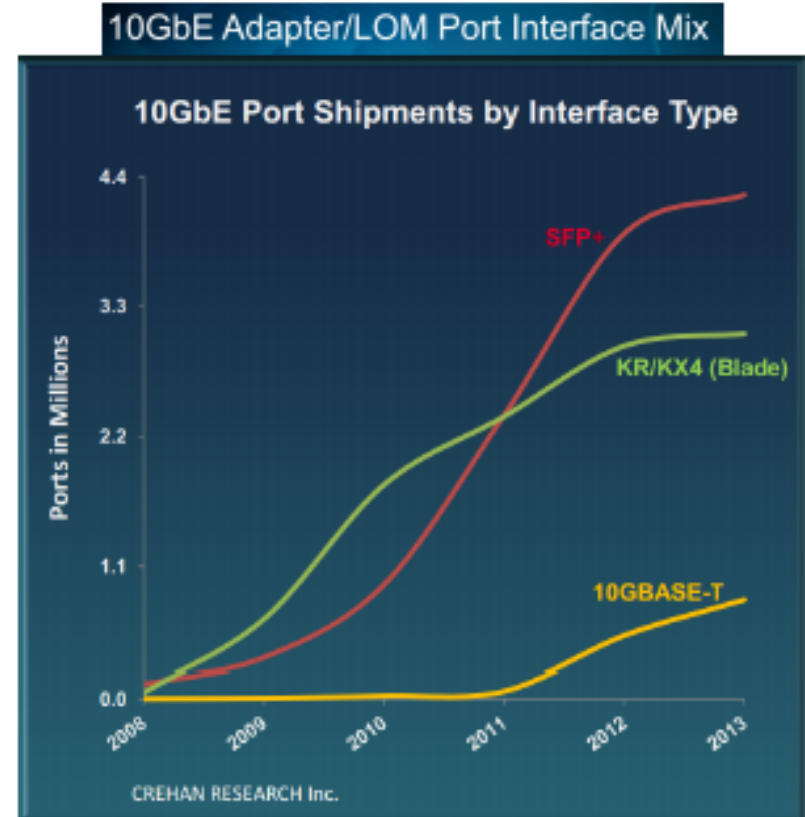
Ethernet Ports using single-lane (10GE)

Data from 10GbE shipments

- 10GbE volume ramp in servers coincided with the availability of single-lane interfaces
- Early adopters (2004-2008) used
 - XAUI-based optics
 - 10GBASE-CX4
 - 10GBASE-KX4
- Single-lane backplane and twinax solutions eclipsed the early-adopter volume starting in 2009

Chart notes

- "Other" category, not shown, went from ~12% in 2008 to <1% in 2013
- SFP+ majority use is twinax, then SR; accurate share data unavailable
- Blade server is mostly KR based upon system configuration. KX4 vs. KR split data unavailable.



Data source: Crehan Research, Inc., Q1'2014

IEEE 802.3 Call For Interest – 25Gb/s Ethernet over a single lane for server interconnect – July 2014 San Diego

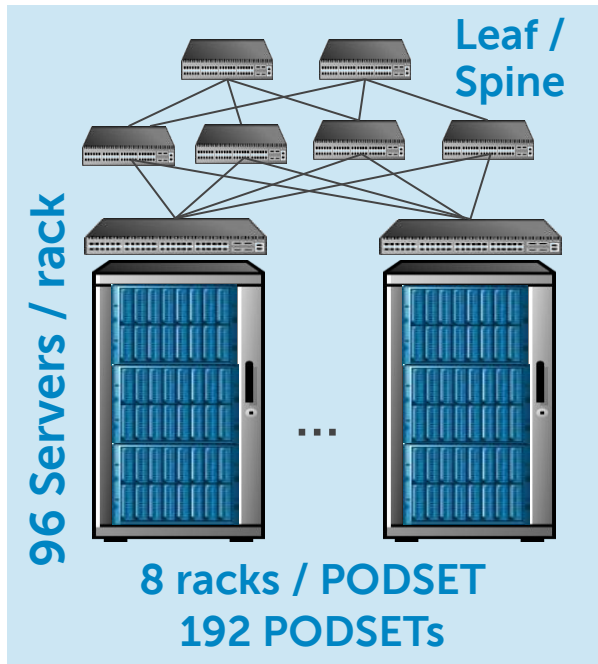
From IEEE 802.3 25GbE CFI presentation, July 2014



Ethernet Ports using single-lane (25GbE)

25GbE Drivers

- Economics drove it, a large market (Cloud DC) pulled it
- Alignment to major serdes / signaling rate - technology reuse, while enabling single lane server I/O



<i>A Data center design</i>	25 GbE	40 GbE
# of Servers	147,456	
# of ToRs	1536	6,144
# 4x25 DAC (breakout)	36,864	n/a
# 40GbE (4X10) DAC (p2p)	n/a	147,456
# of Spine Devices	64	
# of Leaf Devices	768	
#100G Optic links	24,576	

- **25 GbE reduces CAPEX & OPEX!**

Plan for 50Gb/s serial Ethernet

Leading Application: Data center end-point IO

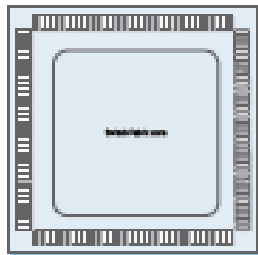
- 50G signaling work driven by IEEE 802.3bs, OIF 56G, is catalyst for development of next-gen 50GbE
- Ethernet speed aligned to major serdes / signaling rate
 - Success with 10GbE
 - A major motivation for 25GbE
 - Why would 50G be any different?
- Servers can fill greater I/O bandwidth as it gets developed.
 - Convergence, virtualization, scaling trends.
 - Servers already using 40GbE, will use 50G (2x25G)
- Plan for 50GbE standardization now



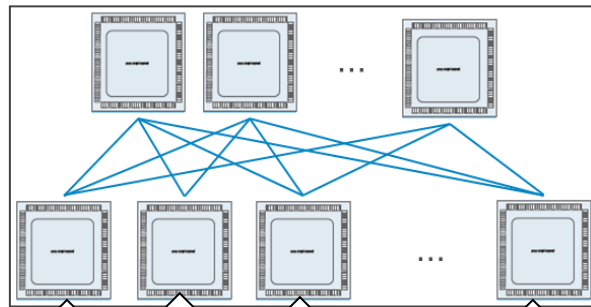
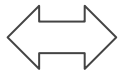
Ethernet Ports using 4x Lanes (40GbE)

40G (4x10G) switches provided the radix to build-out large, flat Data center architectures

- Example: 128 serdes switch ASIC building block

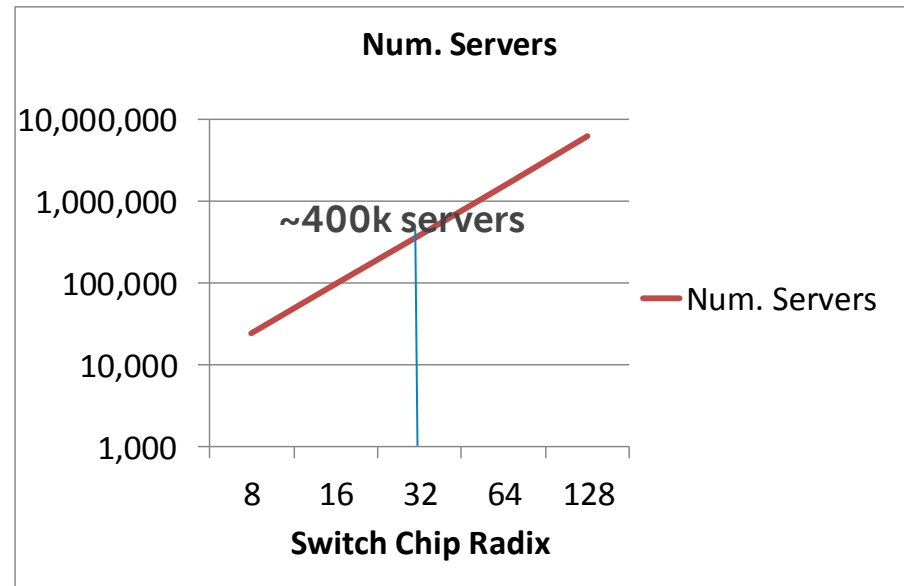


128 x 10GbE, or
32 x 40GbE, or
12 x 100GbE.



Large port count (Spine) switch
(E.g. 288 x 40G)

- Data center scale-out $\sim O(F^2)$; F = switch radix



40G (4x10) provided sufficient switch radix, 100G (10x10) did not.

Ethernet Ports using 4x lanes (40GbE)

40G (4x10) QSFP+ evolved to meet a variety of Data center cabling requirements

- QSFP+ now covers all 4 Optical media Quadrants →
- 4x Lane components have provided compact designs
 - 4x WDM
 - 4x Laser Arrays
 - 4x Modulators
 - 4x Receivers
 - Etc.
- 4x10GE breakout was key!

• QSFP+ coverage

	Duplex	MMF	SMF
		✓ 100m	✓ 2km ✓ 10km ✓ 40km
		✓ 100m ✓ 300m	✓ 500m

Ethernet Ports using 4x lanes (100GbE)

100GE (4x25G) is set to replicate 40GE (4x10) paradigm

- With 25GE servers, 100GE (4x25G) will be preferred Network Port speed
- 100GE (4x25G) switch ASICs will provide the radix needed for large Data centers build-outs
- 100GE QSFP28 will evolve (like 40GE QSFP+) to meet various Data center cabling needs
- 4x25GbE Breakout is key.
- Little to no change in network/cabling architecture over 10G servers / 40GbE networking

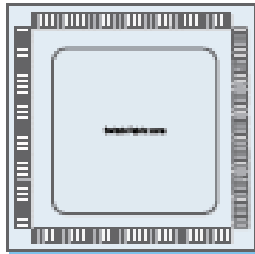


Ethernet Ports using 4x lanes (200GbE)

This trend could continue with 50G serdes / signaling

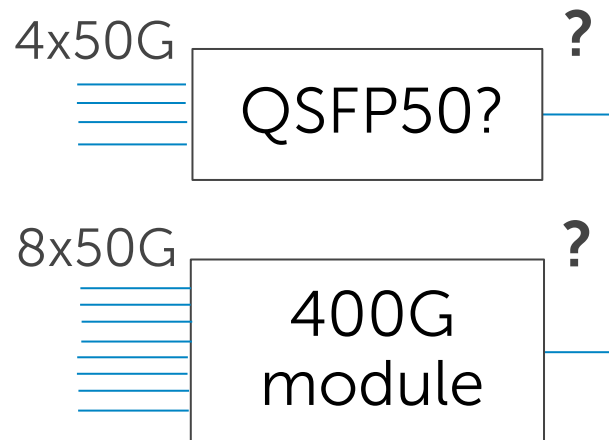
- Strong track record with Quad Modules.
- Smaller challenge for switch chips to maintain radix
- 50G servers and 200GbE Networking can continue using Data center architectures used for 10/40GE, 25/100GE.
- But not an Ethernet MAC rate that is being considered.

50G serdes switch





- **N x 400GbE**
- **2N x 200GbE**

Speed / FF that meets DC needs better?



- QSFP+ coverage

	Duplex	MMF	SMF
		✓ 100m	✓ 2km ✓ 10km ✓ 40km
	✓ 100m ✓ 300m		✓ 500m



Or scale using 100GbE (2x50G)?

- Option 1: 2x50G C2C, Gearbox to CAUI4, use existing 100GbE (4x25G) PMDs and QSFP28.
 - + Continue to use QSFP28 / 4x25G PMDs
 - + Backwards compatibility to systems using 25G serdes.
 - Lower face-plate density than a 200G (4x50G) FF
- Option 2: 2x50G C2C, 2x50G C2M, 2x50G PMDs
 - + Solves face-plate density issue compared to Option 1
 - New set of 100G (2x50G) optical PMDs, some (if not all) will have to go into QSFP28 for inter-op
 - Similar to 40G (2x20G), 2x20G PMDs were not defined.
- Case for 4x50G (200GbE) looks stronger.

Conclusion

- 50G serdes / signaling work in 802.3bs and OIF 56G will act as a catalyst for 50GbE definition
- 50GbE based on a single-lane is a natural follow on to 10GbE and 25GbE as a speed for DC end-points
- 50GbE server IO should make us think very hard about what network port speeds will get used in Data centers
- Switch Radix, module FF, feasibility of optical PMDs, efficient breakout to 50GE, DC scale-out architectures, are major factors in deciding optimum Network Port speed
- Based on past history of Quad lane speeds and modules, 200GbE (4x50G) could be a very compelling Data center network port speed.

